

Aspergillus Genomes and the Aspergillus Cloud

Jane E. Mabey Gilsenan^{1,*}, Graham Atherton¹, Jennifer Bartholomew¹,
Peter F. Giles¹, Teresa K. Attwood^{2,3}, David W. Denning¹ and Paul Bowyer¹

¹School of Medicine, The University Hospital of South Manchester (Wythenshawe), Manchester M23 9LT,

²Faculty of Life Sciences and ³School of Computer Science, The University of Manchester, Oxford Road, Manchester M13 9PT, UK

Received September 14, 2008; Accepted October 18, 2008

ABSTRACT

Aspergillus Genomes is a public resource for viewing annotated genes predicted by various *Aspergillus* sequencing projects. It has arisen from the union of two significant resources: the *Aspergillus*/Aspergillosis website and the Central *Aspergillus* Data REpository (CADRE). The former has primarily served the medical community, providing information about *Aspergillus* and associated diseases to medics, patients and scientists; the latter has focused on the fungal genomic community, providing a central repository for sequences and annotation extracted from *Aspergillus* Genomes. By merging these databases, genomes benefit from extensive cross-linking with medical information to create a unique resource, spanning genomics and clinical aspects of the genus. *Aspergillus Genomes* is accessible from <http://www.aspergillus-genomes.org.uk>.

INTRODUCTION

An important consequence of the impact of global climate change, brought to the fore by recent natural disasters, has been a stimulation of interest in fungal ecology. In the aftermath of Hurricanes Katrina and Rita, for example, residential areas in New Orleans, having remained underwater for weeks, succumbed to rapid mould growth (1–3), facing residents with significant health hazards: exposure to high concentrations of some moulds (e.g. *Aspergillus*, *Cladosporium* and *Trichoderma*) can cause severe and sometimes life-threatening responses. Examination of samples taken from affected homes identified several *Aspergillus* species, *Aspergillus niger* being predominant amongst them. The concentration of this and other species reached levels generally associated with environmental health problems. The paucity of effective drugs available to treat *Aspergillus* infections, coupled with the speed of

diagnosis required for successful treatment, raises concerns for residents and tradesmen who return to restore such areas without taking appropriate protective measures.

Generally, *Aspergillus* is a genus of fungus found worldwide: approximately 250 species descriptions have been published (4,5), several of which are of medical or industrial importance. *Aspergillus fumigatus* is the most common mould pathogen of humans, causing both life-threatening invasive disease of immuno-compromised patients and allergic disease in patients with atopic immune systems (6–8). *Aspergillus nidulans*, an occasional human pathogen, is a model organism that has contributed to our understanding of genetics, gene regulation and cellular biology (9,10), while *A. niger* (11,12) and *A. oryzae* (13) are both used in industrial processes. *Aspergillus flavus* is both a human and plant pathogen, being responsible for a disproportionate number of cutaneous and wound infections in man (14). Several other *Aspergillus* species are known to be significant allergens or to be responsible for mycotoxin production in stored food (15–17).

In light of their relevance to medicine and industry, and the desire to better understand this genus, the genomes of 10 *Aspergilli* have recently been sequenced, seven of which have been annotated in worldwide collaborative efforts. To be of maximum benefit to the community in general, this expanding pool of genomic data (the volume of which is likely to increase with the decreasing cost of sequencing) requires collation and long-term maintenance; more particularly, to benefit research into Aspergillosis and other illnesses caused by this genus, the accumulating genomic data needs to be cross-linked to the numerous *Aspergillus*-related medical reports now available. However, the standard model for genome databases [e.g. SGD (18)] is not appropriate for medically significant organisms, where the dissemination of information, such as patient care, drug resistance or toxins is equally important. Thus, *Aspergillus Genomes* was established in April 2008 to provide links between genomic and medical information and to provide access through one portal.

*To whom correspondence should be addressed. Tel: +44 161 2915908; Fax: +44 161 2915804 Email: jane.gilsenan@manchester.ac.uk

SOURCE DATA AND METHODS

This resource is a joint undertaking by two teams that have, until now, provided online medical and genomic information for different groups within the *Aspergillus* community.

Medical information

The *Aspergillus/Aspergillosis* website (<http://www.aspergillus.org.uk>) was designed to serve the medical and translational research communities, including medical consultants, scientists, patients and their relatives. Offering a wealth of information about *Aspergillus* and the multiple diseases (known collectively as Aspergillosis) it can cause, one section provides advice and a discussion group for patients, while other areas provide medical and species images, educational materials, meeting reports and publications. As part of Aspergillus Genomes, this large collection of material offers a valuable repository from which we can begin to derive much-needed medical insights from genome analyses (Some sections require registration, please refer to supplement 'Note: Access information').

Genomic information

The Central *Aspergillus* Data REpository, CADRE (<http://www.cadre-genomes.org.uk>) (19), was developed primarily to serve the *Aspergillus* genomics community. Its principal role has been to manage genomic data and to offer web-based tools for analysis and visualization of genomic features. These tools offer simple displays for viewing annotation of predicted genes (e.g. function, GO terms, similarity matches) and of their protein products (e.g. family and domain similarity matches), as well as complex displays for viewing genes and other features (e.g. RNA-encoding genes, repeated sequences) in the context of an assembly.

Incorporation of this resource into Aspergillus Genomes provides access to seven annotated genomes, five of which are of particular importance to the *Aspergillus* genetics community (*A. fumigatus* Af293 and A1163, *A. nidulans*, *A. oryzae* and *A. niger*); of these, two are clinical isolates (*A. fumigatus* Af293 and A1163).

CONTENT OF CURRENT RELEASE

Aspergillus Genomes has been implemented using Ensembl version 22 (20), a suite devised for storing annotated eukaryotic genomes. The latest release (June 2008) contains information pertaining to seven genomes.

Aspergillus fumigatus (Af293) 2.2.1

Using whole-genome shotgun (WGS) sequencing (carried out by TIGR) and optical mapping, this project yielded an assembly of 19 supercontigs (~28.8 Mb) mapping onto eight chromosomes. Using automated gene predictions, the University of Manchester coordinated an international manual annotation project, and data for 9926 genes were released in 2005 (21). This genome has since been re-annotated as part of an additional international comparative project led by the J. Craig Venter Institute

(JCVI) and the University of Manchester. Data for 9630 predicted protein-coding genes were publicly released in 2008 (22).

Aspergillus nidulans (FGSC A4) 5.3.1

WGS sequencing of this genome was undertaken by the Broad Institute and publicly released in March 2003. This project yielded an assembly of 17 supercontigs (~30 Mb) assigned to eight linkage groups, a group of unassigned fragments (84 contigs) and 9520 predicted protein-coding genes. Further automated and manual annotation carried out by TIGR (2005) yielded a total of 10 542 genes (23)—using publicly available linkage data (John Clutterbuck, <http://www.gla.ac.uk/ibls/molgen/aspergillus/index.html>) and in-house analyses, we have updated the original Broad assembly to reflect these findings. More contigs have now been mapped, with correct orientation, within each supercontig; the supercontigs have also been orientated correctly within each linkage group.

Aspergillus oryzae (RIB 40) 1.1.1

WGS sequencing of this genome was undertaken by a Japanese Consortium, led by the National Institute of Advanced Industrial Science and Technology (AIST). Using optical mapping, the project yielded an assembly of 22 supercontigs (~37.1 Mb) mapped to eight chromosomes and a group of unassigned fragments. Using automated annotation techniques, 12 074 protein-coding genes were predicted and publicly released in 2005 (24).

Aspergillus niger (CIB 513.88) 1.1.1

Using WGS sequencing and BAC walking (carried out by DSM Food Specialties), this project yielded 19 supercontigs (~33.9 Mb) mapping onto eight chromosomes. Using a base of automated gene predictions, DSM coordinated a European manual annotation project, predicting 14 086 protein-coding genes, and data were released in 2007 (25).

Aspergillus fumigatus (A1163) 1.1.1, *A. clavatus* (NRRL 1) 1.1.1 and *Neosartorya fischeri* (NRRL 181) 1.1.1

These genomes were sequenced using WGS sequencing (carried out by JCVI), as part of an international comparative project led by JCVI and the University of Manchester (2008) (22). The work on *A. fumigatus* yielded 55 scaffolds (~29 Mb), for which 9929 protein-coding genes were predicted; *A. clavatus* yielded 143 unassigned fragments (~27.8 Mb), for which 9121 protein-coding genes were automatically predicted; and *N. fischeri* yielded 976 unassigned fragments (~32.5 Mb), for which 10 406 protein-coding genes were automatically predicted.

DISPLAY AND SEARCH SOFTWARE

Several tools are provided (via CADRE) for viewing genomic data within Aspergillus Genomes. GeneView (Supplementary Figure S1) is the principal data-visualization tool, offering detailed information about a particular gene, including the public locus, the chromosomal location and a short description of the gene. An overview of each predicted transcript is also provided: this consists of

structure information, a list of database cross-references to similar sequences (e.g. SGD) and protein features [e.g. InterPro (26)], and GO terms that have been mapped to the gene. Images are also displayed of each transcript structure and of domains or family signatures mapped to the protein.

In addition to the information provided in GeneView, Transview provides the transcript sequence, which can be marked up to highlight both codons and the peptide sequence. ExonView provides the nucleotide sequence for each exon, as well as the upstream and downstream regions, untranslated regions and introns. ProtView displays further information about a particular protein and, in addition to the GeneView summary it provides the sequence in FASTA format, which can be marked up to highlight exons. ProtView also provides some predicted peptide statistics and information about any matches to family- or domain-based databases [e.g. Pfam (27), PRINTS (28)].

Four search tools, all accessible from the homepage, are available for Aspergillus Genomes: UniSearch, GoogleTM customized search, the Aspergillus Cloud and BLAST. UniSearch, the in-built Ensembl search tool, allows users to search public identifiers assigned to genes (e.g. AFUB_063690 and An14g07380)—it is found on the homepage and throughout the website. A familiar tool is provided in the form of a customized GoogleTM search engine, thus enabling free-text searches on all information pages and genomes within the resource. A more recent addition is ‘cloud’ searching, which is described in more detail below. To allow similarity searches, BLAST can be performed against in-house sequences (i.e. genomic sequence, predicted ORFs and proteins), the results of which are presented in the context of the assembly rather than as stand-alone alignments.

THE ASPERGILLUS CLOUD

The results of UniSearch or GoogleTM searches are provided in the form of lists of pages and documents matching the given term. Although useful in many cases, this type of output does not empower the user to access and search related data. To remedy this, we have introduced the Aspergillus Cloud, a facility that uses ‘cloud’ searching (developed by QuinturaTM) across the medical and genomic information embraced by Aspergillus Genomes. Cloud searching provides a visual means of searching related terms: when a search term is entered, this generates a group (or cloud) of related terms from context, as well as a list of hyperlinks to relevant pages/documents. Furthermore, by exploring the cloud (i.e. by moving a mouse over any term), search results change to reflect relationships with the new highlighted term.

We can demonstrate the Aspergillus Cloud using documented research on point mutations in the cyp51a gene and its correlation to rising triazole drug resistance. By simply entering ‘cyp51a’ into the search box (Figure 1), a number of related terms are displayed, as well as a number of hyperlinks to relevant pages. Several links provide access to pages that are contained within the medical

section of Aspergillus Genomes and refer to point mutations within the gene and drug resistance; one provides access to the relevant gene within the genomic section of Aspergillus Genomes. Medical terms can also be used to explore the Aspergillus Cloud (Figure 2). By entering the term ‘Allergic Bronchopulmonary Aspergillosis’ (ABPA; an allergic reaction to *Aspergillus* spores), opportunities are given to explore articles written about the allergen to which ABPA patients are hypersensitive, as well as annotation provided for an ABPA-related gene.

The Aspergillus Cloud provides a powerful means of exploring Aspergillus data and finding medical reports linked to particular genes that would require several search attempts or would otherwise be missed using conventional search tools.

EXPLORING FEATURES OF INTEREST

The collation of *Aspergillus* data enables us to focus on and embellish features of medical interest. *Aspergillus fumigatus*, for example, possesses over 80 genes encoding allergen proteins and of these 22 have been cloned as IgE binding (29,30). We have documented these allergens within Aspergillus Genomes and provided links to annotated genes (http://www.aspergillus.org.uk/indexhome.htm?secure/sequence_info/genesofinterest/allergens.php). Another area of great interest is secondary metabolites. We have already extracted information on many secondary metabolites (http://www.aspergillus.org.uk/secure/metabolites/list_by_secmet.php) and intend to link this data to annotated genes. We also have begun the process of highlighting secondary metabolite clusters, such as the gliotoxin biosynthetic cluster. Gliotoxin is a toxin produced by *A. fumigatus* and possesses many biological properties, including the ability to act as an antibacterial and antiviral agent as well as an immunosuppressant (31,32). We have provided a view of the gene cluster within the assembly and links to the appropriate annotation (Figure 3).

FUTURE DIRECTIONS

In the short term, we plan to add three more new genomes and to revise a fourth. The new genomes include *A. flavus* NRRL 3357 (TIGR, 2005), *A. terreus* (Broad Institute, 2006) and *A. niger* ATCC 1015 (DOE Joint Genome Institute, 2006); although these genomes have been sequenced with high coverage, only draft sequence and limited annotation are currently publicly available—we will add the sequence data to our collection for comparative work and add annotation as it becomes available. As for the latter, we are currently participating in Eurofungbase, a European project coordinated by Eurofung (<http://www.eurofung.net>), to manually re-annotate *A. nidulans* with the help of experts from 32 laboratories and 13 industrial partners. Our role is to house the corrected gene structures and annotation arising from this project, and to make the data publicly available: we are processing this information in-house, and will release it through CADRE and Aspergillus Genomes.

(a) The Aspergillus Cloud search interface. A search bar at the top contains 'cyp51a'. Below it, a tree diagram shows relationships between terms: mutation, fumigatus, combination, sterol, aspergillus, alpha, cyp51a, gene, demethylase, alteration, identical. A red arrow points from the 'cyp51a' node to a list of search results on the right.

(b) Aspergillus Articles page for record 391. Title: Differences in interactions between azole drugs related to modifications in the 14-alpha sterol demethylase gene (cyp51A) of *Aspergillus fumigatus*. Author: Garcia-Effron G, Mellado E, Gomez-Lopez A, Alcazar-Fuoli L, Cuena-Estrella M, Rodriguez-Tudela JM. Date added to website: 2005-06-07. Abstract: The combined activity of different azole drugs was investigated. Thirty-one *Aspergillus fumigatus* strains were tested, including two cyp51A^{-/-} and one cyp51B^{-/-} gene-knockout strain andazole-susceptible and -resistant strains with different resistance mechanisms. The combination of itraconazole and voriconazole was synergistic for all strains except for those with gene knockouts.

(c) CADRE A. fumigatus (AF293) Gene View. Gene ID: AFUA_4G06890. Genomic Location: 2141884 - 2143502 bp (2.1 Mb) on supercontig 96. Description: At4g06890 (Alt. Locus), 14-alpha sterol demethylase Cyp51A, identical to 14-alpha sterol demethylase (GI|14861413) (*Aspergillus fumigatus*) PMID: 11427550, (1.14.3.). Prediction Method: First-pass computational and manual annotation were carried out via TIGI and the Wellcome Trust Sanger Institute, as TIGI is part of the project. Several programs were also used by the Sanger Institute, as TIGI is part of the project. These include a variety of programs, i.e., NAP, BLASTP, InterProScan, HMM2, ProteinSearch, Picard, SignalP and TMHMM. Second-pass computational and manual annotation were carried out via the JCVI as part of an international comparative project in 2007. Transcript Transcript Neighborhood: Shows the genomic region with exons and introns. Export Data: Export gene data in EMBL, GenBank or FASTA.

Figure 1. Using the Aspergillus Cloud to explore genes. Entering a gene name, such as 'cyp51a', into the search box at the top of the cloud search space (a) produces numerous related search terms immediately below and a list of relevant documents/records at the RHS. The document list provides an extract for each document enabling the user to assess the context of the search term found. By clicking on relevant links, the user can access both medical reports (b) and genomic annotation (c) for a given gene name.

(a) The Aspergillus Cloud search interface. A search bar at the top contains 'allergic bronchopulmonary aspergillosis'. Below it, a tree diagram shows relationships between terms: lung, sinusitis, patient, stable, bronchopulmonary aspergillosis, allergic, asthma, case, lesson. A red arrow points from the 'allergic' node to a list of search results on the right.

(b) Aspergillus Articles page for record 594. Title: Asp f6, an *Aspergillus* allergen specifically recognized by IgE from patients with allergic bronchopulmonary aspergillosis (ABPA). Author: Schweiwacher M, Israel L, Heesemann J, Ebel F. Date added to website: 2006-04-18. Abstract: *Aspergillus fumigatus* is a pathogenic mould causing allergic and invasive respiratory diseases. Allergic bronchopulmonary Aspergillosis (ABPA) is a severe pulmonary complication resulting from hypersensitivity to *A. fumigatus* proteins. *Aspergillus* allergen Asp f6 is recognized by IgE from ABPA patients, but not from sensitized patients without ABPA. There is a difference between the two groups of allergic patients. METHODS: Proteins from hyphae, resting and germinating conidia of *A. fumigatus* were compared by SDS-PAGE. Protein identification was performed using MALDI-TOF mass spectrometry. Recombinant

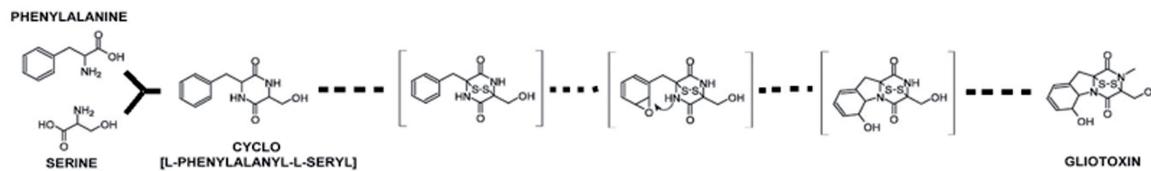
(c) Gene Report for An03g00770. Gene ID: An03g00770. Genomic Location: 30060 - 31181 bp (30.1 Kb) on contig An03c0030. Description: This gene is located in sequence: An03c0030. Prediction Method: First-pass computational annotation was carried out by a European consortium led by DSM Food Specialties. ORFs were chiefly predicted using FGENESH. PEDANT-ProTm was used to analyse each ORF, for which predicted features were manually verified and annotated.

(d) Aspergillus Articles page for record 594. Title: Asp f6, an *Aspergillus* allergen specifically recognized by IgE from patients with allergic bronchopulmonary aspergillosis (ABPA). Author: Schweiwacher M, Israel L, Heesemann J, Ebel F. Date added to website: 2006-04-18. Abstract: *Aspergillus fumigatus* is a pathogenic mould causing allergic and invasive respiratory diseases. Allergic bronchopulmonary Aspergillosis (ABPA) is a severe pulmonary complication resulting from hypersensitivity to *A. fumigatus* proteins. *Aspergillus* allergen Asp f6 is recognized by IgE from ABPA patients, but not from sensitized patients without ABPA. There is a difference between the two groups of allergic patients. METHODS: Proteins from hyphae, resting and germinating conidia of *A. fumigatus* were compared by SDS-PAGE. Protein identification was performed using MALDI-TOF mass spectrometry. Recombinant

Figure 2. Using the Aspergillus Cloud to explore medical information. Entering a medical condition, such as ABPA, into the search box (a), produces several groups of related terms within the search space below. By moving the mouse over each term, new clouds arise temporarily. Clicking on a term will add the associated clouds into the search space and change the focus of the search to include the term (b). In this example, we find a list of documents on the RHS that includes links to an ABPA-related gene (c) and to an article regarding an allergen that is particular to the allergic response of ABPA patients (d).

Cluster 27. Gliotoxin biosynthetic cluster.

Gliotoxin is regarded as one of the most important toxins produced by *A. fumigatus*. You can link to the secondary metabolites and toxins page [here](#) for more information on this compound.



Gliotoxin is synthesised from the amino acids phenylalanine and serine by a nonribosomal peptide synthetase GliP. The resulting L-phenylalanyl-L-seryl compound then appears to undergo a series of oxidative enzymatic transformations before conversion to mature gliotoxin by a methylation step. Evidence for this reaction pathway is limited but comparison with sirodesmin biosynthesis in *Leptosphaeria maculans* has provided valuable insights. (Cramer RA et al. Disruption of a Nonribosomal Peptide Synthetase in *Aspergillus fumigatus* Eliminates Gliotoxin Production. *Eukaryot Cell*. 2006; 5(6): 972–980, Gardiner and Howlett, Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus* FEMS Micro Lett 2006; 248:241-248)

Cluster 27 information:



Cluster 27 map. Click on the image to go to an active cluster 27 map with links to genes in the database.

Cluster 27							
	Gene ID	Start position	Orientation	End position	Length	Protein ID	Annotation
ATCA_00001	ATCA_00001	0	259	259	259	ATCA_00001	Unknown protein
ATCA_00002	ATCA_00002	0	259	259	259	ATCA_00002	Unknown hypothetical protein
ATCA_00003	ATCA_00003	0	259	259	259	ATCA_00003	Unknown hypothetical protein
ATCA_00004	ATCA_00004	0	259	259	259	ATCA_00004	Unknown hypothetical protein
ATCA_00005	ATCA_00005	0	259	259	259	ATCA_00005	Unknown hypothetical protein
ATCA_00006	ATCA_00006	0	259	259	259	ATCA_00006	Unknown hypothetical protein
ATCA_00007	ATCA_00007	0	259	259	259	ATCA_00007	Unknown hypothetical protein
ATCA_00008	ATCA_00008	0	259	259	259	ATCA_00008	Unknown hypothetical protein
ATCA_00009	ATCA_00009	0	259	259	259	ATCA_00009	Unknown hypothetical protein
ATCA_00010	ATCA_00010	0	259	259	259	ATCA_00010	Unknown hypothetical protein
ATCA_00011	ATCA_00011	0	259	259	259	ATCA_00011	Unknown hypothetical protein
ATCA_00012	ATCA_00012	0	259	259	259	ATCA_00012	Unknown hypothetical protein
ATCA_00013	ATCA_00013	0	259	259	259	ATCA_00013	Unknown hypothetical protein
ATCA_00014	ATCA_00014	0	259	259	259	ATCA_00014	Unknown hypothetical protein
ATCA_00015	ATCA_00015	0	259	259	259	ATCA_00015	Unknown hypothetical protein
ATCA_00016	ATCA_00016	0	259	259	259	ATCA_00016	Unknown hypothetical protein
ATCA_00017	ATCA_00017	0	259	259	259	ATCA_00017	Unknown hypothetical protein
ATCA_00018	ATCA_00018	0	259	259	259	ATCA_00018	Unknown hypothetical protein
ATCA_00019	ATCA_00019	0	259	259	259	ATCA_00019	Unknown hypothetical protein
ATCA_00020	ATCA_00020	0	259	259	259	ATCA_00020	Unknown hypothetical protein
ATCA_00021	ATCA_00021	0	259	259	259	ATCA_00021	Unknown hypothetical protein
ATCA_00022	ATCA_00022	0	259	259	259	ATCA_00022	Unknown hypothetical protein
ATCA_00023	ATCA_00023	0	259	259	259	ATCA_00023	Unknown hypothetical protein
ATCA_00024	ATCA_00024	0	259	259	259	ATCA_00024	Unknown hypothetical protein
ATCA_00025	ATCA_00025	0	259	259	259	ATCA_00025	Unknown hypothetical protein
ATCA_00026	ATCA_00026	0	259	259	259	ATCA_00026	Unknown hypothetical protein
ATCA_00027	ATCA_00027	0	259	259	259	ATCA_00027	Unknown hypothetical protein
ATCA_00028	ATCA_00028	0	259	259	259	ATCA_00028	Unknown hypothetical protein
ATCA_00029	ATCA_00029	0	259	259	259	ATCA_00029	Unknown hypothetical protein
ATCA_00030	ATCA_00030	0	259	259	259	ATCA_00030	Unknown hypothetical protein
ATCA_00031	ATCA_00031	0	259	259	259	ATCA_00031	Unknown hypothetical protein
ATCA_00032	ATCA_00032	0	259	259	259	ATCA_00032	Unknown hypothetical protein
ATCA_00033	ATCA_00033	0	259	259	259	ATCA_00033	Unknown hypothetical protein
ATCA_00034	ATCA_00034	0	259	259	259	ATCA_00034	Unknown hypothetical protein
ATCA_00035	ATCA_00035	0	259	259	259	ATCA_00035	Unknown hypothetical protein
ATCA_00036	ATCA_00036	0	259	259	259	ATCA_00036	Unknown hypothetical protein
ATCA_00037	ATCA_00037	0	259	259	259	ATCA_00037	Unknown hypothetical protein
ATCA_00038	ATCA_00038	0	259	259	259	ATCA_00038	Unknown hypothetical protein
ATCA_00039	ATCA_00039	0	259	259	259	ATCA_00039	Unknown hypothetical protein
ATCA_00040	ATCA_00040	0	259	259	259	ATCA_00040	Unknown hypothetical protein
ATCA_00041	ATCA_00041	0	259	259	259	ATCA_00041	Unknown hypothetical protein
ATCA_00042	ATCA_00042	0	259	259	259	ATCA_00042	Unknown hypothetical protein
ATCA_00043	ATCA_00043	0	259	259	259	ATCA_00043	Unknown hypothetical protein
ATCA_00044	ATCA_00044	0	259	259	259	ATCA_00044	Unknown hypothetical protein
ATCA_00045	ATCA_00045	0	259	259	259	ATCA_00045	Unknown hypothetical protein
ATCA_00046	ATCA_00046	0	259	259	259	ATCA_00046	Unknown hypothetical protein
ATCA_00047	ATCA_00047	0	259	259	259	ATCA_00047	Unknown hypothetical protein
ATCA_00048	ATCA_00048	0	259	259	259	ATCA_00048	Unknown hypothetical protein
ATCA_00049	ATCA_00049	0	259	259	259	ATCA_00049	Unknown hypothetical protein
ATCA_00050	ATCA_00050	0	259	259	259	ATCA_00050	Unknown hypothetical protein
ATCA_00051	ATCA_00051	0	259	259	259	ATCA_00051	Unknown hypothetical protein
ATCA_00052	ATCA_00052	0	259	259	259	ATCA_00052	Unknown hypothetical protein
ATCA_00053	ATCA_00053	0	259	259	259	ATCA_00053	Unknown hypothetical protein
ATCA_00054	ATCA_00054	0	259	259	259	ATCA_00054	Unknown hypothetical protein
ATCA_00055	ATCA_00055	0	259	259	259	ATCA_00055	Unknown hypothetical protein
ATCA_00056	ATCA_00056	0	259	259	259	ATCA_00056	Unknown hypothetical protein
ATCA_00057	ATCA_00057	0	259	259	259	ATCA_00057	Unknown hypothetical protein
ATCA_00058	ATCA_00058	0	259	259	259	ATCA_00058	Unknown hypothetical protein
ATCA_00059	ATCA_00059	0	259	259	259	ATCA_00059	Unknown hypothetical protein
ATCA_00060	ATCA_00060	0	259	259	259	ATCA_00060	Unknown hypothetical protein
ATCA_00061	ATCA_00061	0	259	259	259	ATCA_00061	Unknown hypothetical protein
ATCA_00062	ATCA_00062	0	259	259	259	ATCA_00062	Unknown hypothetical protein
ATCA_00063	ATCA_00063	0	259	259	259	ATCA_00063	Unknown hypothetical protein
ATCA_00064	ATCA_00064	0	259	259	259	ATCA_00064	Unknown hypothetical protein
ATCA_00065	ATCA_00065	0	259	259	259	ATCA_00065	Unknown hypothetical protein
ATCA_00066	ATCA_00066	0	259	259	259	ATCA_00066	Unknown hypothetical protein
ATCA_00067	ATCA_00067	0	259	259	259	ATCA_00067	Unknown hypothetical protein
ATCA_00068	ATCA_00068	0	259	259	259	ATCA_00068	Unknown hypothetical protein
ATCA_00069	ATCA_00069	0	259	259	259	ATCA_00069	Unknown hypothetical protein
ATCA_00070	ATCA_00070	0	259	259	259	ATCA_00070	Unknown hypothetical protein
ATCA_00071	ATCA_00071	0	259	259	259	ATCA_00071	Unknown hypothetical protein
ATCA_00072	ATCA_00072	0	259	259	259	ATCA_00072	Unknown hypothetical protein
ATCA_00073	ATCA_00073	0	259	259	259	ATCA_00073	Unknown hypothetical protein
ATCA_00074	ATCA_00074	0	259	259	259	ATCA_00074	Unknown hypothetical protein
ATCA_00075	ATCA_00075	0	259	259	259	ATCA_00075	Unknown hypothetical protein
ATCA_00076	ATCA_00076	0	259	259	259	ATCA_00076	Unknown hypothetical protein
ATCA_00077	ATCA_00077	0	259	259	259	ATCA_00077	Unknown hypothetical protein
ATCA_00078	ATCA_00078	0	259	259	259	ATCA_00078	Unknown hypothetical protein
ATCA_00079	ATCA_00079	0	259	259	259	ATCA_00079	Unknown hypothetical protein
ATCA_00080	ATCA_00080	0	259	259	259	ATCA_00080	Unknown hypothetical protein
ATCA_00081	ATCA_00081	0	259	259	259	ATCA_00081	Unknown hypothetical protein
ATCA_00082	ATCA_00082	0	259	259	259	ATCA_00082	Unknown hypothetical protein
ATCA_00083	ATCA_00083	0	259	259	259	ATCA_00083	Unknown hypothetical protein
ATCA_00084	ATCA_00084	0	259	259	259	ATCA_00084	Unknown hypothetical protein
ATCA_00085	ATCA_00085	0	259	259	259	ATCA_00085	Unknown hypothetical protein
ATCA_00086	ATCA_00086	0	259	259	259	ATCA_00086	Unknown hypothetical protein
ATCA_00087	ATCA_00087	0	259	259	259	ATCA_00087	Unknown hypothetical protein
ATCA_00088	ATCA_00088	0	259	259	259	ATCA_00088	Unknown hypothetical protein
ATCA_00089	ATCA_00089	0	259	259	259	ATCA_00089	Unknown hypothetical protein
ATCA_00090	ATCA_00090	0	259	259	259	ATCA_00090	Unknown hypothetical protein
ATCA_00091	ATCA_00091	0	259	259	259	ATCA_00091	Unknown hypothetical protein
ATCA_00092	ATCA_00092	0	259	259	259	ATCA_00092	Unknown hypothetical protein
ATCA_00093	ATCA_00093	0	259	259	259	ATCA_00093	Unknown hypothetical protein
ATCA_00094	ATCA_00094	0	259	259	259	ATCA_00094	Unknown hypothetical protein
ATCA_00095	ATCA_00095	0	259	259	259	ATCA_00095	Unknown hypothetical protein
ATCA_00096	ATCA_00096	0	259	259	259	ATCA_00096	Unknown hypothetical protein
ATCA_00097	ATCA_00097	0	259	259	259	ATCA_00097	Unknown hypothetical protein
ATCA_00098	ATCA_00098	0	259	259	259	ATCA_00098	Unknown hypothetical protein
ATCA_00099	ATCA_00099	0	259	259	259	ATCA_00099	Unknown hypothetical protein
ATCA_00100	ATCA_00100	0	259	259	259	ATCA_00100	Unknown hypothetical protein

Cluster 27 Gene Table. Click on the image to go to the cluster 27 gene table with links to genes in the database.

In the longer term, we will continue to expand the section dealing with areas of medical interest. Explicit links and clustering of such data enables rapid assimilation of essential information. In addition, to support in-house work and provide users with value-added data, we will expand our comparative analyses and include resultant data within Aspergillus Genomes. Access to similar data is currently facilitated by the Broad *Aspergillus* Comparative Database (http://www.broad.mit.edu/annotation/genome/aspergillus_group/MultiHome.html); however, in contrast with Aspergillus Genomes, which places much attention on the underlying data, the Broad site focuses on function. By providing up-to-date primary data, we hope to yield consistent secondary data, as well as relevant links to current medical information.

CONCLUSIONS

The *Aspergilli* garner interest owing to their medical and industrial significance, but changes in the environment are making the interest in this genus and its effects on the human condition more prominent. It is therefore timely to begin the process of marrying medical and

ACKNOWLEDGEMENTS

We wish to thank all those who have contributed to Aspergillus Genomes by their involvement in the sequencing and annotation projects. Interactions with the *A. nidulans* research community are facilitated by workshops organized by Eurofung (Eurofungbase project—EC 1274195).

FUNDING

The Fungal Research Trust.

Conflict of interest statement: None declared.

REFERENCES

1. Chew,G.L., Wilson,J., Rabito,F.A., Grimsley,F., Iqbal,S., Reponen,T., Muilenberg,M.L., Thorne,P.S., Dearborn,D.G., Morley,R.L. *et al.* (2006) Mold and endotoxin levels in the aftermath of Hurricane Katrina: a pilot project of homes in New Orleans undergoing renovation. *Environ. Health Perspect.*, **114**, 1883–1889.
2. Solomon,G.M., Hjelmroos-Koski,M., Rotkin-Ellman,M. and Hammond,S.K. (2006) Airborne mold and endotoxin concentrations in New Orleans, Louisiana, after flooding, October through November 2005. *Environ. Health Perspect.*, **114**, 1381–1386.
3. Rao,C.Y., Riggs,M.A., Chew,G.L., Muilenberg,M.L., Thorne,P.S., Van Sickle,D., Dunn,K.H. and Brown,C. (2007) Characterization of airborne molds, endotoxins, and glucans in homes in New Orleans after Hurricanes Katrina and Rita. *Appl. Environ. Microbiol.*, **73**, 1630–1634.
4. Pitt,J.I., Samson,R.A. and Frisvad,J.C. (2000) Accepted species and their synonyms in the Trichocomaceae. In Samson,R.A. and Pitt,J.I. (eds), *Integration of Modern Taxonomic Methods for Penicillium and Aspergillus classification*, Harwood Publishers, Amsterdam, pp. 9–49.
5. Geiser,D.M., Klich,M.A., Frisvad,J.C., Peterson,S.W., Varga,J. and Samson,R.A. (2007) The current status of species recognition and identification in Aspergillus. *Stud. Mycol.*, **59**, 1–10.
6. Denning,D.W. (1998) Invasive Aspergillosis. *Clin. Infect. Dis.*, **26**, 781–803.
7. Latgé,J.-P. (1999) *Aspergillus fumigatus* and Aspergillosis. *Clin. Microbiol. Rev.*, **12**, 310–350.
8. Marr,K.A., Patterson,T. and Denning,D. (2002) Aspergillosis. Pathogenesis, clinical manifestations, and therapy. *Infect. Dis. Clin. North Am.*, **16**, 875–894.
9. Pontecorvo,G., Roper,J.A., Hemmons,L.M., MacDonald,K.D. and Button,A.W.J. (1953) The genetics of *Aspergillus nidulans*. *Adv. Genet.*, **5**, 141–239.
10. Martinelli,S.D. and Kinghorn,J.R. (eds) (1994) *Aspergillus: 50 years on*. Elsevier, Amsterdam.
11. Bennett,J.W. and Klich,M.A. (1999) Aspergillus. In Flickinger,M.C. and Drew,S.W. (eds), *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis, and Bioseparation*. John Wiley and Sons, Inc., New York, pp. 213–220.
12. Ruijter,G.J.G., Kubicek,C.P. and Visser,J. (2002) Production of organic acids by the fungi. In Osiewacz,H.D. (ed.), *Industrial Applications*, Vol. X, Springer-Verlag, Berlin, pp. 213–230.
13. Nout,M.J. and Aidoo,K.E. (2002) Asian fungal fermented food. In Osiewacz,H.D. (ed.), *Industrial Applications*, Vol. X, Springer-Verlag, Berlin, pp. 23–47.
14. Hedayati,M.T., Pasqualotto,A.C., Warn,P.A., Bowyer,P. and Denning,D.W. (2007) Aspergillus flavus: human pathogen, allergen and mycotoxin producer. *Microbiology*, **153**, 1677–1692.
15. Blyth,W., Grant,I.W., Blackadder,E.S. and Greenberg,M. (1977) Fungal antigens as a source of sensitization and respiratory disease in Scottish maltworkers. *Clin. Allergy*, **7**, 549–562.
16. Jarvis,J.Q. and Morey,P.R. (2001) Allergic respiratory disease and fungal remediation in a building in a subtropical climate. *Appl. Occup. Environ. Hyg.*, **16**, 380–388.
17. Pitt,J.I. (2000) Toxicogenic fungi: which are important? *Med. Mycol.*, **38**(Suppl. 1), 17–22.
18. Hirschman,J.E., Balakrishnan,R., Christie,K.R., Costanzo,M.C., Dwight,S.S., Engel,S.R., Fisk,D.G., Hong,E.L., Livstone,M.S., Nash,R. *et al.* (2006) Genome snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res.*, **34**, D442–D445.
19. Mabey,J.E., Anderson,M.J., Giles,P.F., Miller,C.J., Attwood,T.K., Paton,N.W., Bornberg-Bauer,E., Robson,G.D., Oliver,S.G. and Denning,D.W. (2004) CADRE: the Central *Aspergillus* Data REpository. *Nucleic Acids Res.*, **32**, D401–D405.
20. Flieck,P., Aken,B.L., Beal,K., Ballester,B., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cunningham,F., Cutts,T. *et al.* (2008) Ensembl 2008. *Nucleic Acids Res.*, **36**, D707–D714.
21. Nierman,W.C., Pain,A., Anderson,M.J., Wortman,J.R., Kim,H.S., Arroyo,J., Berriman,M., Abe,K., Archer,D.B., Bermejo,C. *et al.* (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature*, **438**, 1151–1156.
22. Fedorova,N.D., Khaldi,N., Joardar,V.S., Maiti,R., Amedeo,P., Anderson,M.J., Crabtree,J., Silva,J.C., Badger,J.H., Albarraq,A. *et al.* (2008) Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genet.*, **4**, e1000046.
23. Galagan,J.E., Calvo,S.E., Cuomo,C., Ma,L.J., Wortman,J.R., Batzoglou,S., Lee,S.I., Basturkmen,M., Spevak,C.C., Clutterbuck,J. *et al.* (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, **438**, 1105–1115.
24. Machida,M., Asai,K., Sano,M., Tanaka,T., Kumagai,T., Terai,G., Kusumoto,K., Arima,T., Akita,O., Kashiwagi,Y. *et al.* (2005) Genome sequencing and analysis of *Aspergillus oryzae*. *Nature*, **438**, 1157–1161.
25. Pel,H.J., de Winde,J.H., Archer,D.B., Dyer,P.S., Hofmann,G., Schaap,P.J., Turner,G., de Vries,R.P., Albang,R., Albermann,K. *et al.* (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat. Biotechnol.*, **25**, 221–231.
26. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Buillard,V., Cerutti,L., Copley,R. *et al.* (2007) New developments in the InterPro database. *Nucleic Acids Res.*, **35**, D224–D228.
27. Finn,R.D., Mistry,J., Schuster-Bockler,B., Griffiths-Jones,S., Hollich,V., Lassmann,T., Moxon,S., Marshall,M., Khanna,A., Durbin,R. *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.*, **34**, D247–D251.
28. Attwood,T.K., Bradley,P., Flower,D.R., Gaulton,A., Maudling,N., Mitchell,A.L., Moulton,G., Nordle,A., Paine,K., Taylor,P. *et al.* (2003) PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.*, **31**, 400–402.
29. Bowyer,P., Fraczek,M. and Denning,D.W. (2006) Comparative genomics of fungal allergens and epitopes shows widespread distribution of closely related allergen and epitope orthologues. *BMC Genomics*, **7**, 251.
30. Crameri,R., Limacher,A., Weichel,M., Glaser,A.G., Zeller,S. and Rhyner,C. (2006) Structural aspects and clinical relevance of *Aspergillus fumigatus* antigens/allergens. *Med. Mycol.*, **44**, S261–S267.
31. Cramer,R.A. Jr., Gamesik,M.P., Brooking,R.M., Najvar,L.K., Kirkpatrick,W.R., Patterson,T.F., Balibar,C.J., Graybill,J.R., Perfect,J.R., Abraham,S.N. *et al.* (2006) Disruption of a nonribosomal peptide synthetase in *Aspergillus fumigatus* eliminates gliotoxin production. *Eukaryot Cell*, **5**, 972–980.
32. Gardiner,D.M. and Howlett,B.J. (2005) Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*. *FEMS Microbiol. Lett.*, **248**, 241–248.